Hands-on Scikit-Learn for Machine Learning Applications

Data Science Fundamentals with Python

David Paper

Apress®

Hands-on Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python

David Paper
Logan, UT, USA

ISBN-13 (pbk): 978-1-4842-5372-4 ISBN-13 (electronic): 978-1-4842-5373-1
https://doi.org/10.1007/978-1-4842-5373-1

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Jonathan Gennick
Development Editor: Laura Berendson
Coordinating Editor: Jill Balzano

Cover image designed by Freepik (www.freepik.com)

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail rights@apress.com, or visit http://www.apress.com/rights-permissions.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/9781484253724. For more detailed information, please visit http://www.apress.com/source-code.

Printed on acid-free paper

For my mother, brothers, and friends.

Table of Contents

About the Author

Dr. David Paper is a professor at Utah State University in the Management Information Systems department. He is the author of two books – *Web Programming for Business: PHP Object-Oriented Programming with Oracle* and *Data Science Fundamentals for Python and MongoDB*. He has over 70 publications in refereed journals such as *Organizational Research Methods*, *Communications of the ACM, Information & Management*, *Information Resource Management Journal*, *Communications of the AIS*, *Journal of Information Technology Case and Application Research*, and *Long Range Planning*. He has also served on several editorial boards in various capacities, including associate editor. Besides growing up in family businesses, Dr. Paper has worked for Texas Instruments, DLS, Inc., and the Phoenix Small Business Administration. He has performed IS consulting work for IBM, AT&T, Octel, Utah Department of Transportation, and the Space Dynamics Laboratory. Dr. Paper's teaching and research interests include data science, machine learning, process reengineering, object-oriented programming, and change management.

About the Technical Reviewer

Jojo Moolayil is an artificial intelligence, deep learning, machine learning, and decision science professional and published author of three books: *Smarter Decisions – The Intersection of Internet of Things and Decision Science*, *Learn Keras for Deep Neural Networks*, and *Applied Supervised Learning with R*. He has worked with industry leaders on several high-impact and critical data science and machine learning projects across multiple verticals. He is currently associated with Amazon Web Services as a research scientist – AI.

Jojo was born and raised in Pune, India, and graduated from the University of Pune with a major in Information Technology Engineering. He started his career with Mu Sigma Inc., the world's largest pure-play analytics provider, and worked with the leaders of many Fortune 50 clients. He later worked with Flutura – an IoT analytics start-up – and GE, the pioneer and leader in Industrial AI.

He currently resides in Vancouver, BC. Apart from authoring books on deep learning, decision science, and IoT, Jojo has also been a technical reviewer for various books on the same subject with Apress and Packt publications. He is an active Data Science tutor and maintains a blog at `http://blog.jojomoolayil.com`.

- Jojo's personal web site: `www.jojomoolayil.com`
- Business e-mail: `mail@jojomoolayil.com`

Introduction

We apply the popular Scikit-Learn library to demonstrate machine learning exercises with Python code to help readers solve machine learning problems. The book is designed for those with intermediate programming skills and some experience with machine learning algorithms. We focus on application of the algorithms rather than theory. So, readers should read about the theory online or from other sources if appropriate. The reader should also be willing to spend a lot of time working through the code examples because they are pretty deep. But, the effort will pay off because the examples are intended to help the reader tackle complex problems.

The book is organized into eight chapters. Chapter 1 introduces the topic of machine learning, Anaconda, and Scikit-Learn. Chapters 2 and 3 introduce algorithmic classification. Chapter 2 classifies simple data sets and Chapter 3 classifies complex ones. Chapter 4 introduces predictive modeling with regression. Chapters 5 and 6 introduce classification tuning. Chapter 5 tunes simple data sets and Chapter 6 tunes complex ones. Chapter 7 introduces predictive modeling regression tuning. Chapter 8 puts all knowledge together to review and present findings in a holistic manner.

Download this book's example data by clicking the Download source code button found on the book's catalog page at https://www.apress.com/us/book/9781484253724.